

How Do Washington's Graduation Tests Measure Up?

**A Comparison of the 2003
10th Grade Washington Assessment
of Student Learning with High
School Graduation Exams
from Other States**

About Achieve, Inc.

Created by the nation's governors and business leaders, Achieve, Inc., is a bipartisan, non-profit organization that helps states raise academic standards, improve assessments and strengthen accountability to prepare all young people for postsecondary education, work and citizenship. Achieve has helped nearly half the states benchmark their standards and tests against the best examples in this country and abroad and work in partnership to improve teaching and learning. Achieve serves as a significant national voice for quality in standards-based reform and regularly convenes governors, CEOs and other influential leaders at National Education Summits and other gatherings to sustain support for higher standards and achievement for all of America's schoolchildren.

Copyright © 2005 Achieve, Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information or storage retrieval system, without permission from Achieve, Inc.

Editorial assistance and design: KSA-Plus Communications, Inc.



How Do Washington's Graduation Tests Measure Up?

A Comparison of the
2003 10th Grade Washington
Assessment of Student Learning
with High School Graduation
Exams from Other States

Table of Contents

Executive Summary	1
I. Background	5
II. How does the WASL compare with other state graduation exams?	7
Reading	8
<i>Content</i>	
Chart 1: Distribution of points by content (8)	
Chart 2: Distribution of points by content: Basic Comprehension (9)	
<i>Item Type</i>	
Chart 3: Distribution of points by item type (10)	
<i>Reading Passage Genre</i>	
Chart 4: Distribution of points by reading passage genre (11)	
<i>Approximate Grade Level</i>	
Chart 5: Distribution of points by ACT level (12)	
Table 1: Distribution of points by ACT level (12)	
<i>Cognitive Demand</i>	
Chart 6: Distribution of points by level of cognitive demand (14)	
<i>Reading Passage Demand</i>	
Chart 7: Distribution of points by reading passage demand (15)	
Table 2: Distribution of points by reading passage demand (15)	
<i>Overall Rigor</i>	
Chart 8: Average rigor of state tests based on Reading Rigor Index (17)	
Writing	18
Mathematics	19
<i>Item Type</i>	
Chart 9: Distribution of points by item type (19)	
<i>Content</i>	
Chart 10: Distribution of points by content (20)	
Table 3: Distribution of points by content (20)	
Table 4: Distribution of points by content: Number (21)	
<i>Algebra</i>	
Chart 11: Distribution of points by content: Algebra (22)	
<i>Geometry</i>	
Table 5: Distribution of points by content: Geometry/Measurement (23)	
<i>Approximate Grade Level</i>	
Chart 12: Test content on the International Grade Placement Scale (24)	
<i>Cognitive Demand</i>	
Chart 13: Distribution of points by level of cognitive demand (26)	
Table 6: Distribution of points by level of cognitive demand (26)	
Chart 14: Distribution of points by level of cognitive demand (27)	
III. How do the performance levels on the WASL compare with those of other states?	28
Performance Levels on the Reading Test	29
Chart 15: Difficulty of average passing scenario at “passing” cut score (ACT scale) (30)	
Performance Levels on the Writing Test	31
Performance Levels on the Mathematics Test	32
Chart 16: Average difficulty of mathematical content at “passing” cut score (International Grade Placement Scale) (32)	
Chart 17: Average cognitive demand of mathematics tests at “passing” cut score (33)	
IV. Conclusion	36
Appendix: Summary of Methodology	37

Executive Summary

In June of 2004, Achieve issued a study comparing the graduation exams in six states — Florida, Maryland, Massachusetts, New Jersey, Ohio and Texas. The study, *Do High School Graduation Exams Measure Up?*, compared the content and rigor of the exams and the cut scores students need to achieve to pass the tests. Achieve launched this study to provide educators, policy-makers and the public with a clearer picture of what high school graduation exams measure and how difficult they are to pass.

After releasing this report, Achieve was asked by the Washington State Academic Achievement and Accountability Commission, the Office of the Superintendent of Public Instruction, and the Partnership for Learning to compare the 2003 10th grade Washington Assessment of Student Learning (WASL) with the six states' exams using the same methodology from the initial study. Because the states that participated in the larger study together enroll nearly a quarter of the nation's high school students, they provide an ideal point of comparison for Washington as it works to improve the WASL over time.

Findings for Washington

After a detailed analysis of the mathematics, reading and writing exams in Washington, Achieve reached conclusions that were similar to those in our initial study. *First*, it is perfectly reasonable to expect high school graduates to pass these tests — they are not overly demanding. *Second*, the exams will need to be strengthened over time to better measure the knowledge and skills high school graduates need to succeed in the real world. *Third*, Washington should not rely exclusively on these tests to measure everything that matters in a young person's education. Like all states, Washington will need to develop over time a more comprehensive set of measures beyond on-demand graduation tests.

The WASL Is Not Overly Demanding

As with the tests across the other six states Achieve has studied, the 2003 WASL in mathematics, reading and writing do not present unreasonable expectations for high school graduates. On the contrary, the tests cover material that most students study by early in their high school careers. Given where the bar is set, it is perfectly reasonable for Washington to require students to pass these exams to earn a high school diploma.

- **The questions on the WASL reflect material that most students study by early in their high school careers.** In mathematics, the WASL places a heavier emphasis on pre-algebra and basic geometry and measurement concepts than on concepts associated with Algebra I and later high school geometry. In English language arts, the WASL focuses on important read-

ing comprehension skills but does not address the more advanced reading skills students will need to succeed in college and the new economy. In addition, the reading passages tend to be less challenging than those of the other states in the 6-state study.

- **The “cut scores” required to pass the tests reflect modest expectations.** To pass the mathematics test, Washington students need to successfully answer questions that, on average, cover material students in most other countries study in 7th or 8th grade. To pass the reading test, students need to successfully answer questions that ACT considers more appropriate for the test it gives to 8th and 9th graders than for its college admissions test. These findings are consistent across other states as well.
- **The tests measure only a fraction of the knowledge and skills that colleges and employers say are essential.** Similar to the tests in the initial study, the WASL in mathematics and English language arts measures some of the skills essential for college and workplace success, but a significant number of those skills go largely unmeasured. The skills that do get measured are fundamental; students cannot succeed without them. But the large gap between these tests and the real-world expectations of colleges and employers suggests that the current tests are not strong measures of college- and workplace-readiness.

Washington’s Writing Test Is Strong

In Achieve’s analysis of Washington’s 2003 assessments, we found that, compared with the other states, the writing test is exemplary. By requiring students to pass that test to graduate, Washington is placing more value on student writing than any of the states in the earlier study, which is commendable given how important strong writing skills are to students’ success in college and careers.

The WASL Should Be Strengthened Over Time

The set of exit exams reviewed by Achieve in 2004 are considerably more challenging than the exams these states once used. However, the Achieve analysis reveals that the bar needs to be raised even higher over time. Achieve recommends that Washington:

- **Emphasize more challenging content.** In mathematics, Washington should increase the rigor of the algebra items on its test and limit the emphasis on number concepts; in reading, Washington should increase the percentage of items that measure upper high school level content and increase the sophistication of the reading passages.

- **Ask more challenging questions.** In mathematics, a majority of the WASL items tap basic skills — e.g., doing routine procedures. Washington should work over time to ensure that a larger percentage of assessment items measure higher-level skills, such as mathematical reasoning. In reading, the items are more balanced, but do not tap the highest level of cognitive demand.
- **Phase in higher cut scores over time.** In addition to increasing the cognitive demand of WASL items, Washington can raise the rigor of the tests over time by raising the score required for passing. Texas is using this approach with its new graduation exam. This strategy will work only if a test has enough range in what it measures, so that a higher score actually reflects more advanced knowledge and skills. If a higher cut score simply means that students must answer more of the same kinds of items correctly, rather than items tapping more advanced concepts and skills, it is not very meaningful to raise the cut score.

Graduation Exams Cannot Measure Everything That Matters

Basic fairness requires that students have multiple opportunities to take high school exit exams, so it is reasonable for states to begin to administer the tests in the 10th or 11th grades. Ultimately, however, it is important for 12th grade students in Washington — and across the country — to be able to do 12th grade work, not just pass a 10th or 11th grade test. Over time, Washington will need to develop a more comprehensive set of measures beyond on-demand graduation tests. For example, the state could develop 12th grade assessments that are well aligned to college and workplace knowledge and skills or, alternatively, end-of-course exams for subjects such as Algebra II or upper-level English that are beyond the range of the exit exams. Rather than attaching high stakes to these tests, the scores might be factored into course grades or included on high school transcripts. This would provide valuable information that postsecondary institutions and employers could use in making admissions, placement or hiring decisions.

Washington also will need to look beyond large-scale assessments because, as critical as they are, they cannot measure everything that matters in a young person's education. The ability to make effective oral arguments and conduct significant research projects are considered essential skills by both employers and postsecondary educators, but these skills are very difficult to assess on a paper-and-pencil test. Washington should work with local districts to develop ways to incorporate research projects and oral examinations into instructional programs and to establish rigorous, systematic criteria for evaluating them across the state.

Conclusion

Achieve launched its original 2004 study to help answer critical questions about the expectations states are setting for their high school graduates through the use of exit exams: Do the tests reflect material that students should be familiar with by the time they complete high school? Is it reasonable to expect all students to pass these tests before they graduate? If they pass these tests, does it mean students are ready for their next steps in life? In Washington and the other six states, we found that the tests do indeed set a “floor” for students that states can responsibly defend as a graduation requirement, but do not effectively measure the higher-level skills that truly constitute “readiness” for college and the world of work.

In states like Washington, where the exit exams are being debated, Achieve strongly encourages policymakers not to lower expectations or delay implementation of stakes. If Washington stays the course while ratcheting up the level of demand of these exams, and makes the necessary investments to improve teaching and learning, it undoubtedly will find that students will rise to the challenge. As sufficient numbers of students pass these tests, Washington should continue to raise the floor to reflect the demands students will face in postsecondary education and the 21st-century workplace.

I. Background

In June of 2004, Achieve published a study comparing the graduation exams in six states — Florida, Maryland, Massachusetts, New Jersey, Ohio and Texas. The study, *Do High School Graduation Exams Measure Up?*, compared the content and rigor of both the tests and the scores that students needed to achieve in order to pass those tests.

After releasing the report, Achieve was asked by the Washington State Academic Achievement and Accountability Commission, the Office of the Superintendent of Public Instruction, and the Partnership for Learning to conduct a similar study comparing the 2003 10th grade Washington Assessment of Student Learning (WASL) to the six states' exams using the same methodology from the larger study. In October 2004, Achieve submitted a summary report intended for discussion at the Commission's October 11th meeting. The summary report was designed to help guide decisions the Commission would be making in the fall of 2004 and to provide information to OSPI that could help improve the WASL over time. The fuller report submitted here is meant to provide the Commission, OSPI and the Partnership for Learning with additional data and greater detail than was included in the October summary report.

Why Achieve Launched the Study of Graduation Exams

High school graduation exams are in place in nearly half the states, and more than half the nation's high school students have to pass them to earn a diploma. More rigorous than an earlier generation of minimum competency tests initiated in the 1980s, these tests are an important part of the decade-long movement to raise standards and improve achievement in the U.S. They also have become a lightning rod for public debate.

The attention exit exams have received is understandable and deserved. They are the most public example of states holding students directly accountable for reaching higher standards. For the most part, however, the public debate over high school exit exams has gone on without vital information about how high a hurdle they actually set in front of high school students.

Achieve launched its 2004 study to provide educators, policymakers and the public with a clearer picture of what high school graduation exams measure and how difficult they are to pass. As a group, the states that participated in the study enroll nearly a quarter of the nation's high school students, making this group of states an ideal point of comparison for Washington as it considers its options for making the WASL part of the graduation requirement.

Methodology

The foundation of Achieve's analysis was a thorough description, grounded in several dimensions, of each test item. Two reviewers trained to use coding schemes for each dimension examined each question and coded it for each dimension. These reviewers worked independently and reconciled any differences in their judgments before final characterizations were assigned to each question. With these detailed analyses of each question, Achieve was able to aggregate the descriptions to build an overall picture of each test, which allowed for cross-state comparisons.

One dimension examined was the content the tests measure to determine what students need to know to pass them. In mathematics, for example, Achieve wanted to know how much algebra appears on the test, and what kind of algebra it is. In this analysis of content, two independently devised benchmarks proved useful, particularly in estimating the grade level of particular content. In mathematics, an international scale created as part of the Third International Mathematics and Science Study (TIMSS) was used. In English, a scale adapted from one used by ACT, Inc. to describe questions on its college preparatory and admissions tests was used.

Another important dimension considered was the complexity of the performance or cognitive demand of each question — e.g., what each question asks students to do with their knowledge in reading, writing and mathematics. In reading, for example, students can be asked to simply recall information from a text — a relatively low level skill — or they can be asked to perform a more complex task such as comparing imagery across different passages.

In addition, the complexity of the reading passages also figured into the analysis, as it is the interaction of cognitive demand and the difficulty of a passage that establishes the rigor of these tests. To address this dynamic, Achieve developed a Reading Rigor Index to rate items.

Finally, this analysis explored what it takes for students to pass each state test and how those expectations compare across states. Achieve and experts from Michigan State University devised a statistical approach to allow cut scores from different states' tests to be compared on the TIMSS and ACT scales. Using this approach, Achieve was able to identify those questions that students who scored at the cut score answered correctly and to determine the content and demand of those items. This helped us paint an overall picture of how challenging each test was to pass relative to the others.

For more information about the methodology used in this analysis, see appendix, page 37.

II. How does the WASL compare with other state graduation exams?

Washington and the other six states have made different policy choices about the timing of their exit exams. Washington, Florida, Ohio and Massachusetts each give their tests for the first time to 10th graders; New Jersey and Texas give their exit exams in the 11th grade; Maryland has created end-of-course exams, with the English exam given as early as the end of 9th grade.

These states also are at different points in the rollout of the assessments. In Florida, Massachusetts and New Jersey the tests already count for high school students, while in Maryland, Ohio, Texas and Washington they will count in the future (see following table).

FL	MD	MA	NJ	OH	TX	WA
TEST						
Florida Comprehensive Assessment Test	High School Assessments	Massachusetts Comprehensive Assessment System	High School Proficiency Assessment	Ohio Graduation Tests	Texas Assessment of Knowledge and Skills	Washington Assessment of Student Learning
GRADE FIRST GIVEN						
10th	End of Course	10th	11th	10th	11th	10th
YEAR FIRST GIVEN						
1998	2001	1998	2002	2003	2003	2001
REPLACED ANOTHER EXIT TEST						
Yes	Yes	No	Yes	Yes	Yes	No
SUBJECTS TESTED FOR GRADUATION REQUIREMENTS						
Reading; mathematics	English I; algebra/data analysis; biology; government	English language arts; mathematics	Reading/writing; mathematics	Reading; mathematics; science; social studies (writing: in development)	English language arts; mathematics; science; social studies	Reading; writing; mathematics
FIRST GRADUATING CLASS REQUIRED TO PASS						
2003	2009*	2003	2003	2007	2005	2008
OPPORTUNITIES FOR STUDENTS WHO HAVE NOT PASSED TO RETAKE TESTS						
Yes	Yes	Yes	Yes	Yes	Yes	Yes
OTHER POLICIES RELATED TO STAKES						
Classes of 2003, 2004 permitted to substitute results on college admissions test.	Proposal is being considered to allow students to fail one of four tests and still graduate with cumulative score across tests.	Appeals process uses statistical comparison of GPAs in subject area courses of passing and non-passing students.	State has alternative, performance-based assessment given and scored locally. Sixteen percent of class of 2003 statewide and up to 50 percent in some districts used this route to graduate.	State law allows students to fail one of five tests and still graduate if score is close to passing mark and GPA in subject is at least 2.5.	Passing score for first two graduating classes was lower than eventual passing mark.	

* State Board of Education approval pending

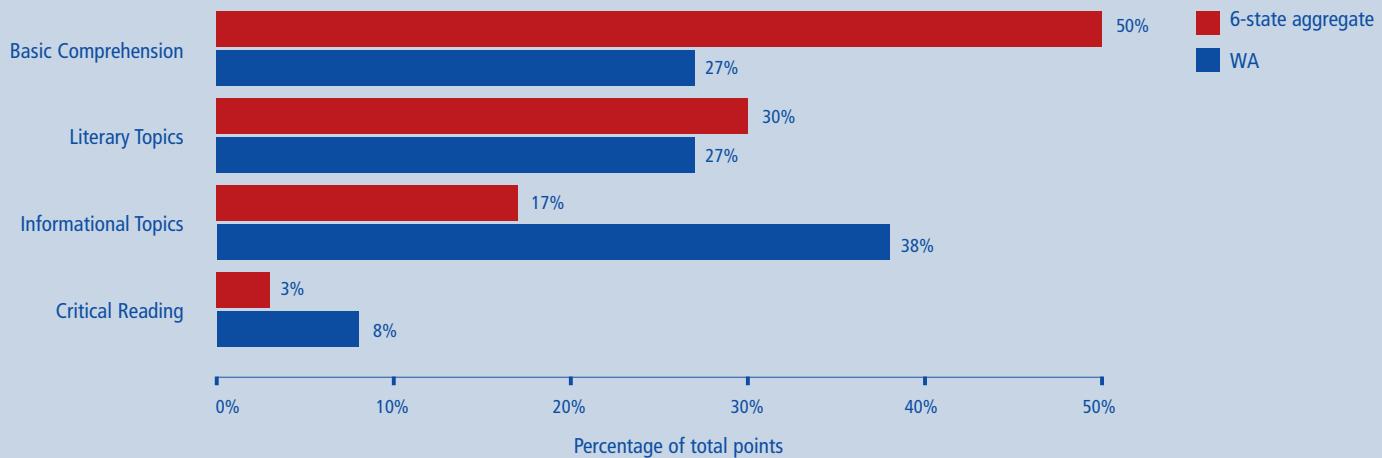
Reading

The WASL emphasizes more advanced content — such as informational topics and critical reading — than other states' tests.

Achieve's initial study showed that state reading tests have one thing in common: They pay greater attention to basic content and less attention to more advanced content. Fifty percent of the total points on the average of the six states' assessments are devoted to basic reading comprehension (e.g., vocabulary; general comprehension of a word, phrase or paragraph; and understanding the main idea or theme of a reading passage). This is not the case on the WASL, where only 27 percent of points are attributed to basic comprehension and the remaining 73 percent are associated with more advanced content, such as literary and informational topics and critical reading.

Furthermore, on the WASL, the emphasis on the characteristics of informational text is higher than other states. Whereas across the six other states, the average percentage of items addressing informational topics is 17 percent, on the WASL a full 38 percent measure these topics. The WASL also devotes a higher percentage of points (8 percent) — albeit a small percentage — to more advanced critical-reading skills, including discerning fact from opinion and faulty from logical reasoning. These are skills that college faculty and frontline managers in a variety of industries agree are essential to success in higher education or on the job.

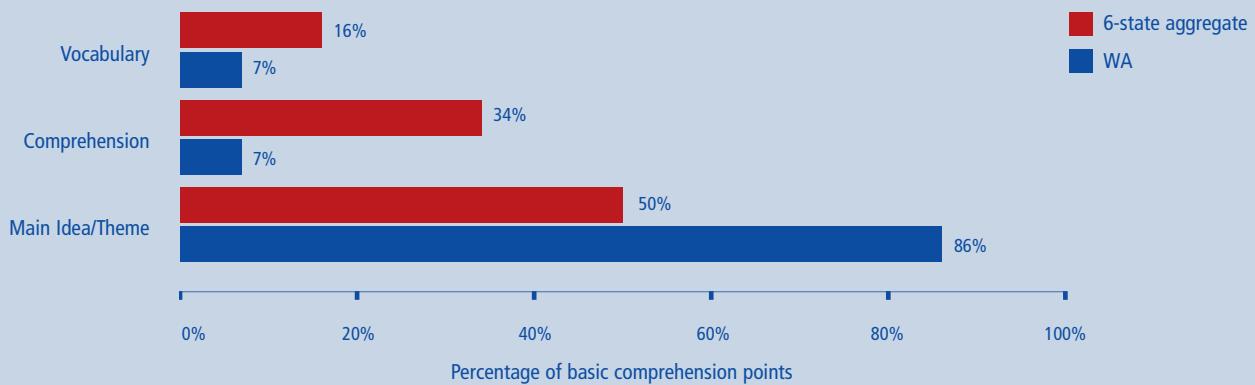
Chart 1: Distribution of points by content



NOTE: *Totals may not equal 100 percent due to rounding.*

A more detailed breakdown of the content categories (see Chart 2) reveals that of the WASL points that are attributed to basic comprehension (e.g., vocabulary, comprehension and main idea/theme items), a large proportion are associated with items that test students' ability to comprehend the main idea or theme of *an entire text*, rather than simply of a word, phrase, or paragraph. This sets them apart from other states, which have a greater number of basic comprehension items that focus on local, rather than global, comprehension. In addition, on the average, other states devote a greater percentage of points to vocabulary items.

Chart 2: Distribution of points by content: Basic comprehension

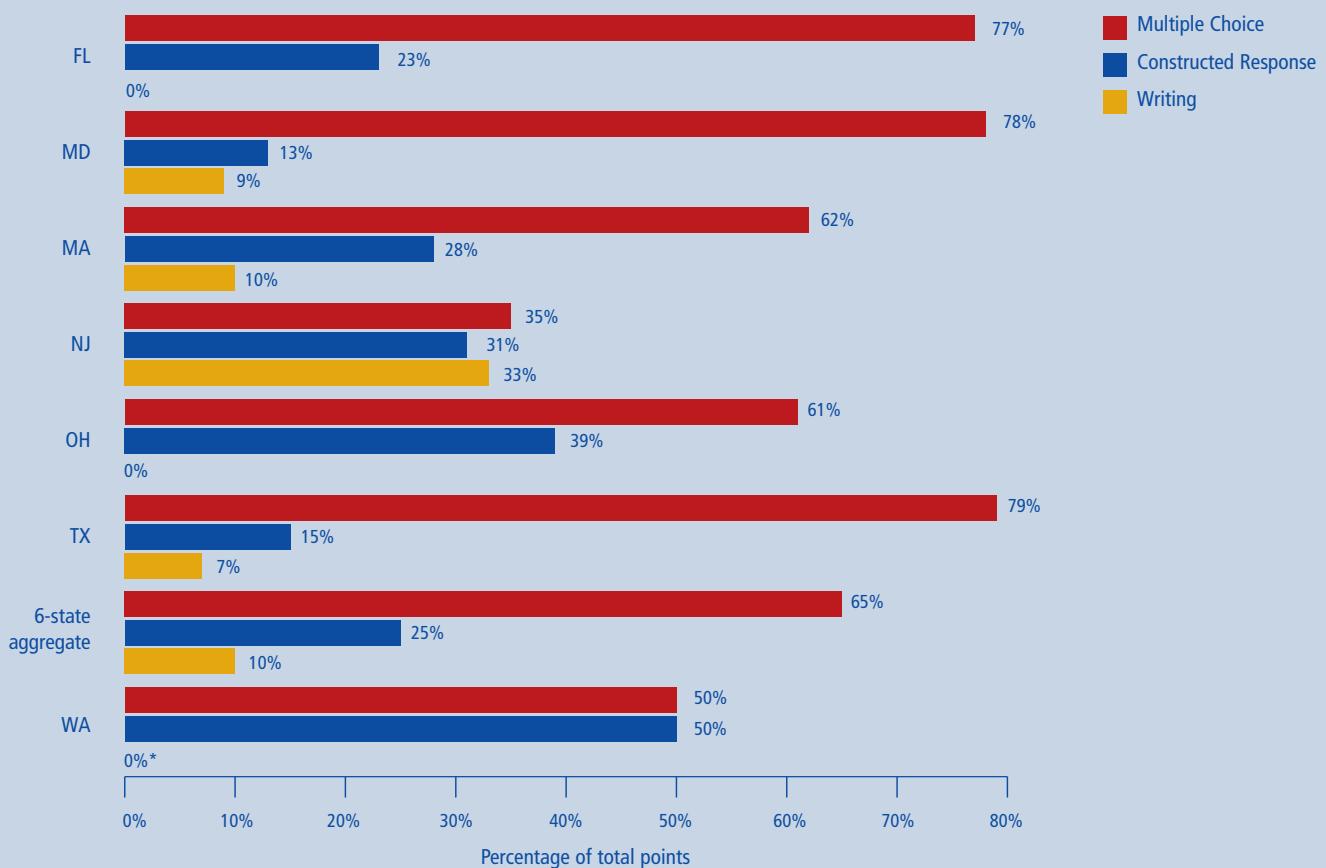


NOTE: *Totals may not equal 100 percent due to rounding.*

The WASL includes more constructed-response items than other states' tests.

The WASL presents an even balance between multiple-choice and constructed-response items that is not evident on the other tests. Of all tests examined, the WASL includes the highest percentage of constructed-response items. This is notable, as constructed response items are often associated with tasks that require more cognitively challenging knowledge and skills.

Chart 3: Distribution of points by item type



NOTE: *Totals may not equal 100 percent due to rounding.*

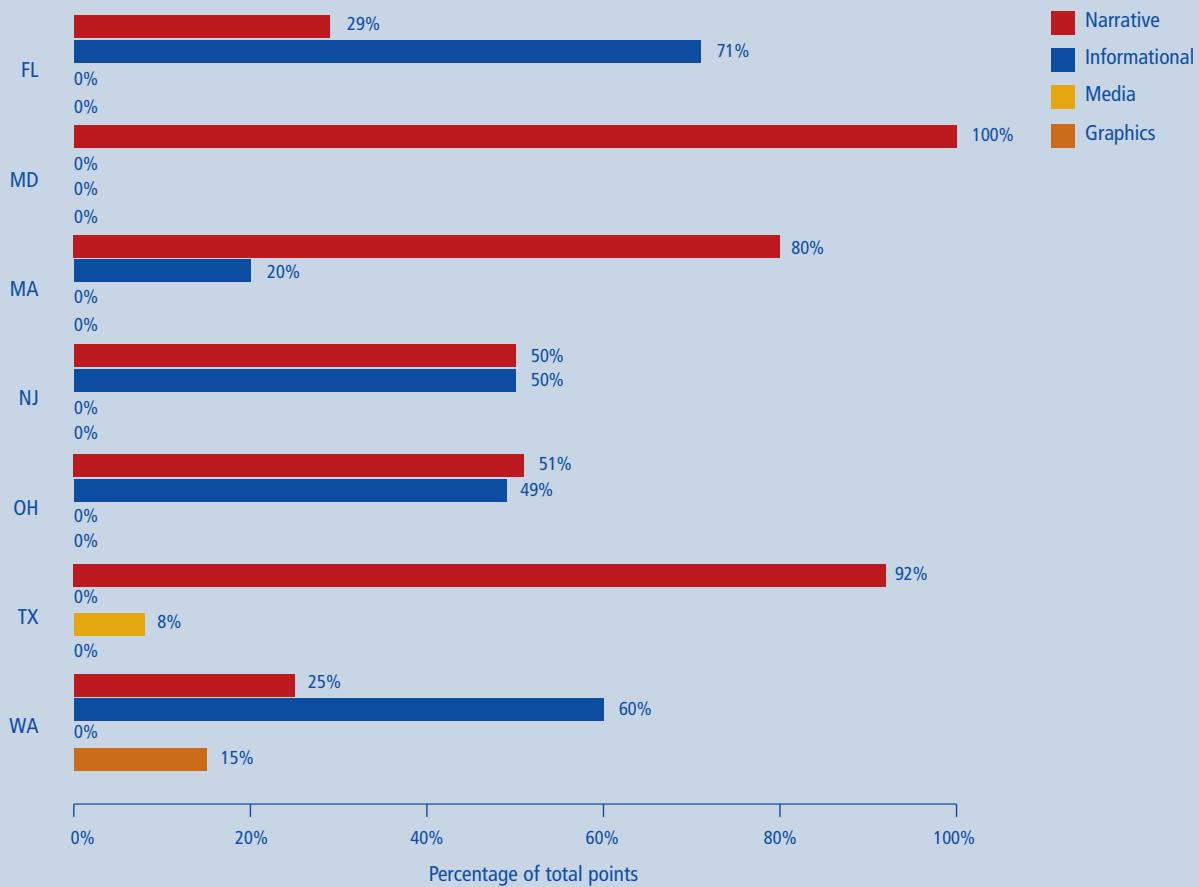
*Washington tests writing separately.

The WASL includes a variety of genres of reading passages, attributing more points to informational text than most other states.

As Achieve's American Diploma Project research found, employers and college professors stress the importance of high school graduates being able to read and interpret a wide range of informational materials, such as periodicals, memoranda, technical writing and intricate charts and graphs. In addition, the current 12th grade NAEP Reading Framework requires that 60 percent of the reading passages on its assessment are informational and 40 percent are literary.

Against this backdrop, it is significant that the WASL prioritizes interpreting informational text by attributing 60 percent of its points to these items. None of the states, with the exception of Florida, examined in the original study emphasized these materials on their exams to the extent that Washington does. In addition, the 15 percent of the WASL points focus on comprehending graphic representations, whereas no other state devotes any specific items to this topic.

Chart 4: Distribution of points by reading passage genre

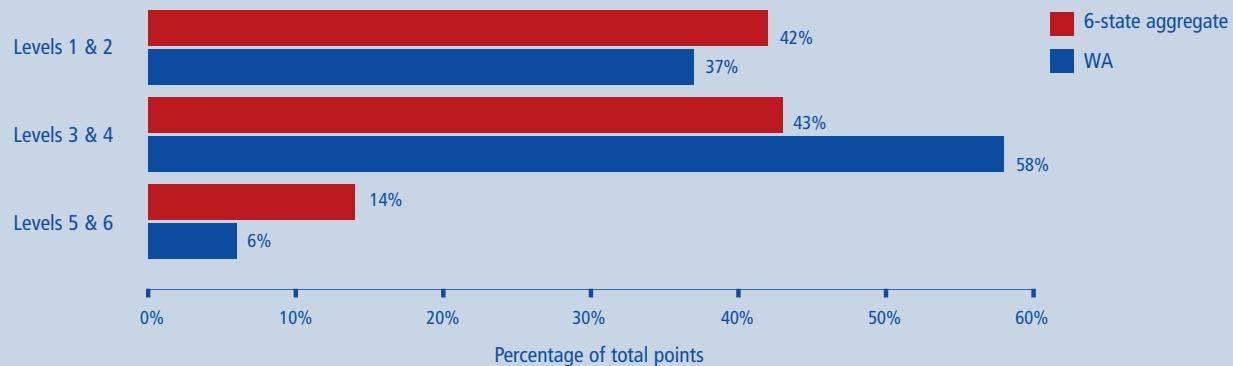


NOTE: *Totals may not equal 100 percent due to rounding.*

The approximate grade level of state reading tests, including the WASL, is late middle school to early high school.

To gauge the approximate grade level of the content on the state exit exams in English language arts, Achieve used an index based on one created by ACT, Inc., to guide the development of assessments given to students as early as the 8th grade. ACT has established six levels to differentiate the content and skills that are measured on its reading tests: Levels 1 through 4 cover skills found on ACT's EXPLORE test given in the 8th and 9th grades; ACT's PLAN test, which is given in the 10th grade, includes test items from Levels 1 through 5; and the ACT Assessment — which students take in the 11th and 12th grades, and which colleges use in admissions, course placement and guidance decisions — incorporates items from Levels 1 through 6.

Chart 5: Distribution of points by ACT level



NOTE: *Totals may not equal 100 percent due to rounding.*

Table 1: Distribution of points by ACT level

Level	ACT EXPLORE (8th and 9th grades)	ACT PLAN (10th grade)	ACT Assessment (11th and 12th grades)	FL	MD	MA	NJ	OH	TX	6-state aggregate	WA
1	10–20%	5–15%	5–15%	10%	7%	24%	3%	16%	3%	11%	27%
2	20–30%			42%	21%	24%	22%	51%	11%	31%	10%
3	30–40%	20–30%	10–20%	19%	11%	20%	19%	27%	24%	21%	33%
4	15–25%	20–30%	20–30%	19%	32%	26%	25%	6%	38%	23%	25%
5	0%	25–35%	25–35%	8%	29%	6%	31%	0%	24%	14%	6%
6	0%	0%	20–30%	2%	0%	0%	0%	0%	0%	0%	0%

As is clear from Table 1, none of the six state tests in our original study approaches the level of demand of the ACT college admissions test. On the contrary, the vast majority of points (86 percent) from the six state tests are tied to ACT Levels 1-4. Thus, when looked at in the aggregate, the level of demand across the six tests most closely resembles that of the ACT EXPLORE test — which is given to students in 8th and 9th grades.

Similarly, the WASL's profile most closely reflects that of the ACT's 8th/9th grade EXPLORE test. However, the WASL has an even smaller percentage of points at Level 5 and 6 (6 percent) than the aggregate of the six states (14 percent) and a higher percentage of Level 1 items than any of the other six states. Across all the tests, the average difficulty of the WASL falls right in the middle, with Texas, Maryland and New Jersey coming out higher and Florida, Massachusetts and Ohio lower.

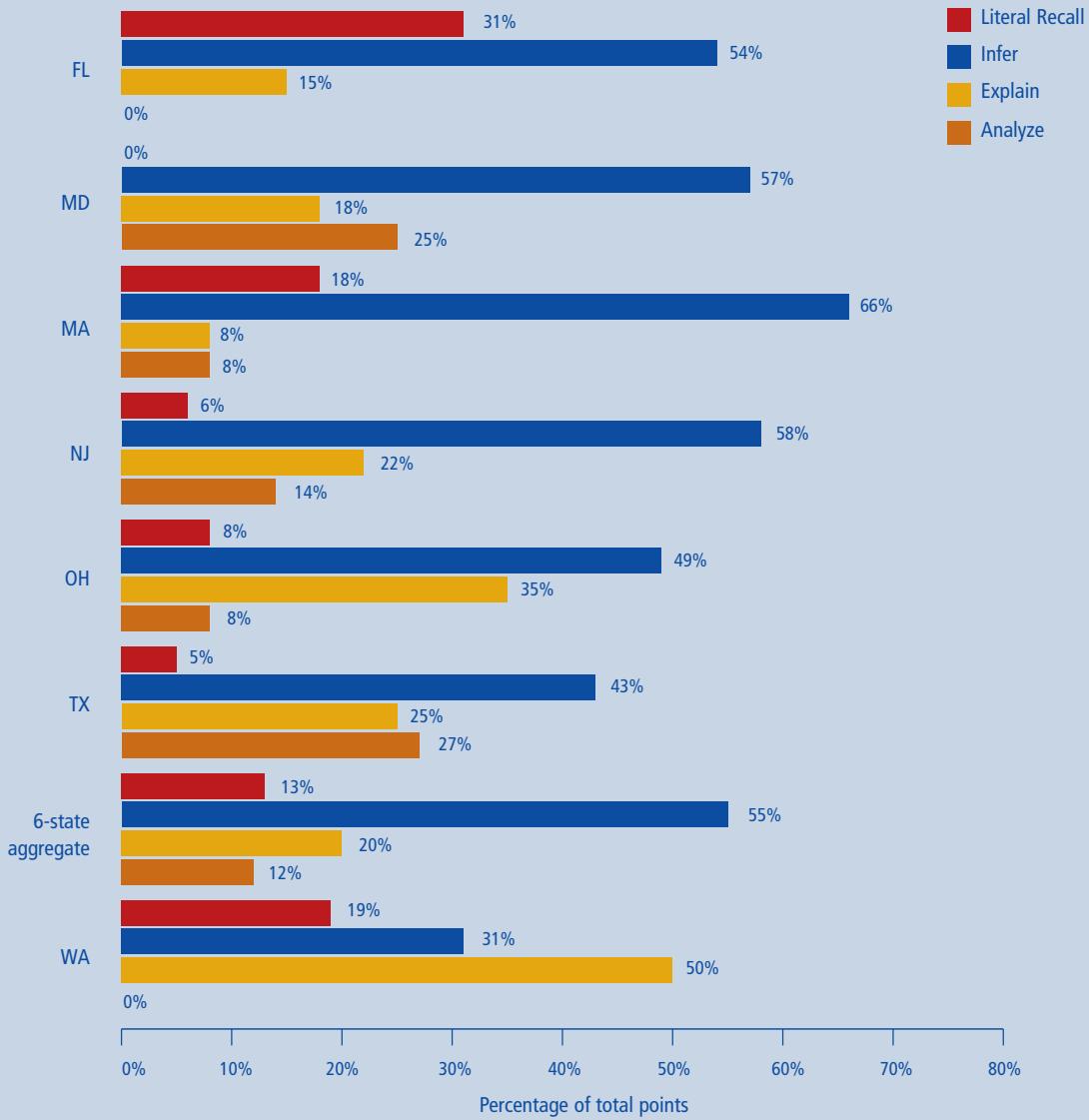
The WASL reading assessment has, on average, a similar level of demand to the other states, but the test includes fewer items at the highest levels of cognitive demand.

A majority of the points across the tests from the original six states are devoted to questions that tap lower-level reading comprehension skills. For example, 68 percent of the points on the tests are associated with skills Achieve considers basic, such as literal recall (13 percent) and inference (55 percent). Twenty percent of the points are associated with questions requiring students to explain — e.g. to provide details to support their answers — and only 12 percent of the total reading points across the six states focus on analysis, which is regarded as the most demanding performance and is exhibited by expert readers.

In comparison, the WASL ties significantly fewer points (50 percent) to basic comprehension skills — literal recall (19 percent) and inference (31 percent) skills. The WASL's remaining points (50 percent) are devoted to explaining, far more than the average (20 percent) of the other states. However, the WASL does not have any items requiring analysis, whereas, on average, across the 6 states in the original study, 12 percent of the points are devoted to this level. Although some question stems in the constructed response items may appear to require some analysis, the scoring guides for these items tend to reward only the citing of examples, not an analysis of the text. This lack of analytical requirements may be a result of the fact that Washington has fewer items that address narrative reading passages relative to the other states studied. Achieve has found that on large-scale state assessments items calling for analysis tend to

address narrative reading passages more frequently than informational reading passages. However, this need not be the case. College preparatory tests such as the ACT Assessment and SAT Reasoning Test include items at this higher level of cognitive demand for both narrative and informational texts.

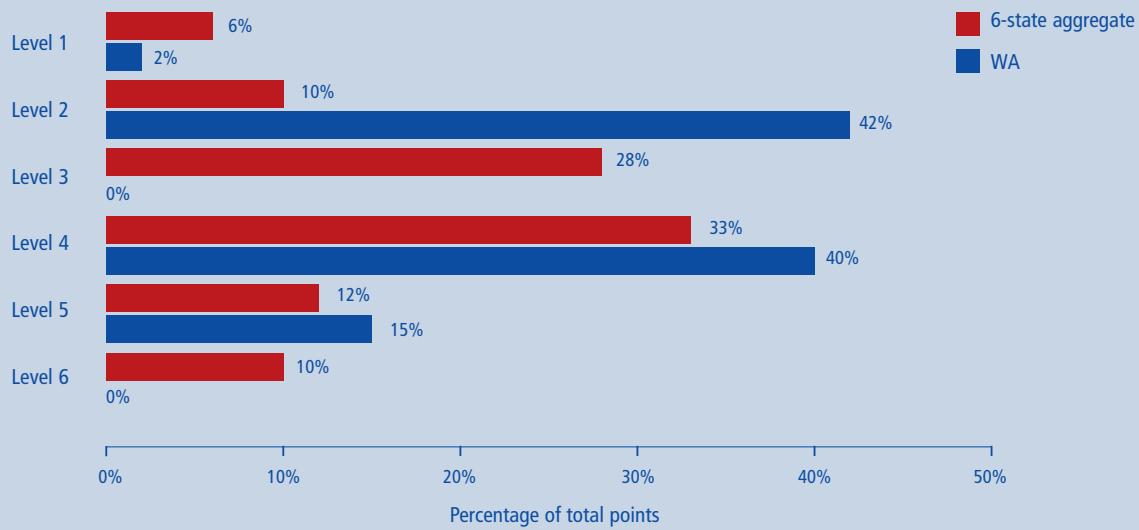
Chart 6: Distribution of points by level of cognitive demand



The reading passages on the WASL are generally less demanding than those of other states' tests; the WASL does not include passages at the highest level of high school reading demand.

To judge the complexity of reading passages, Achieve's reading experts created a six-point scale describing texts from the relatively simple to the very complex. The levels are based on such characteristics as the specialization of the vocabulary, the predictability of text structures or organization, the complexity of the syntax, the level of abstractness, the familiarity of the topic, and the number of concepts introduced in the passage. Level 1 represents upper-elementary reading, Levels 2 and 3 represent middle school reading, Level 4 represents early-stage high school reading, and Levels 5 and 6 represent later-stage high school reading.

Chart 7: Distribution of points by reading passage demand



NOTE: Totals may not equal 100 percent due to rounding.

Table 2: Distribution of points by reading passage demand

Level	FL	MD	MA	NJ	OH	TX	6-state aggregate	WA	
1	0%	0%	8%	0%	14%	8%	6%	2%	
2	0%	22%	0%	0%	37%	0%	10%	42%	
3	37%	37%	22%	50%	25%	0%	28%	0%	
4	63%	41%	20%	0%	24%	51%	33%	40%	
5	0%	0%	33%	0%	0%	41%	12%	15%	
6	0%	0%	16%	50%	0%	0%	10%	0%	

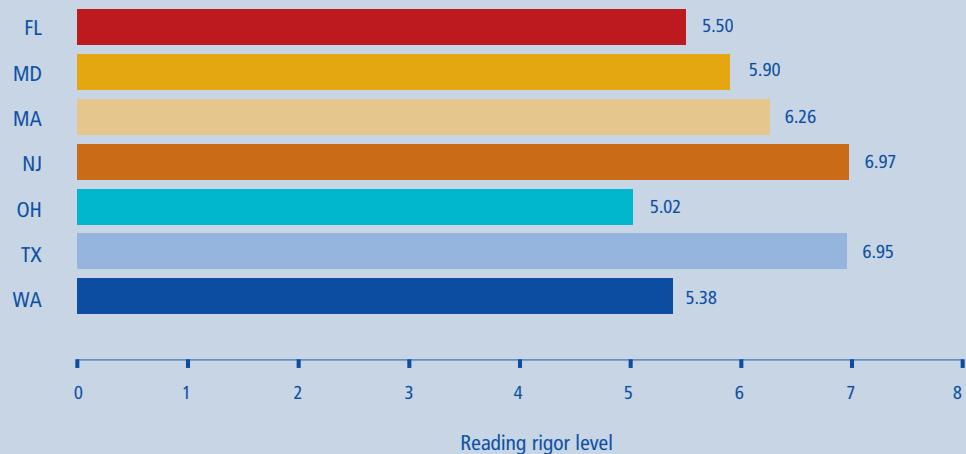
In the original study, across the six states, the majority of points were attributed to reading passages in the middle of this range. Points on the WASL also are distributed across the reading levels, with the majority attributed to level 4 and 5 passages. However, the WASL devotes more points to items assessing passages at the lower levels of rigor than do the other states. Only 16 percent of the points across the six state tests were attributed to passages at the lower two levels of demand, which is appropriate given that these tests are high school graduation exams. In contrast, 44 percent of WASL's points are connected to passages at the two lowest levels of demand. Furthermore, as is the case with most of the states in Achieve's graduation exam study, the WASL does not include any Level 6 reading passages, which are associated with later high school.

The overall rigor of the WASL reading test is below that of most of the other six states.

The difficulty of a reading test is determined not only by the complexity of the reading passages but also by the cognitive demand of the questions about those passages. To capture this important interplay, Achieve developed a Reading Rigor Index (RRI) that combines the cognitive challenge level of an item with the difficulty level of the passage that the item targets. (Note: Cut scores are not factored into the RRI. See appendix for more information on the RRI.)

Based on this scale, the WASL appears to be somewhat less rigorous than most of the other tests, largely because the reading passages are not as challenging. The New Jersey and Texas tests are the most rigorous, followed by Maryland and Massachusetts. The WASL comes close to the Florida test in terms of reading rigor and is more rigorous than Ohio's test. It is worth noting that the two most rigorous tests — Texas and New Jersey — are given in the 11th grade, whereas the rest are 10th grade tests except for Maryland's, which is end-of-course.

Chart 8: Average rigor of state tests based on Reading Rigor Index



Writing

WASL is strong on writing.

Washington's approach to assessing writing on the WASL is as sophisticated as the best of the other states Achieve studied. The WASL writing test is made up exclusively of on-demand writing tasks, with no multiple-choice writing or language questions.

In the original study, four states — Maryland, Massachusetts, New Jersey and Texas — assessed writing in some form on their exit exams. Florida and Ohio plan to include it in the future. The four states that include writing have chosen different approaches to measuring this skill. Two states, New Jersey and Massachusetts, mainly require students to write essays to demonstrate their ability to write in on-demand situations. Maryland, and to a lesser degree Texas, rely on indirect writing measures (multiple-choice items) to assess grammar, punctuation, and editing and revision skills, as well as requiring a direct writing sample.

Washington also sets itself apart from the other states by requiring students to pass the writing assessment to graduate. In three of the other four states, a student's score on the writing items becomes part of her total English language arts score, and better performance in reading can compensate for poor performance in writing. Only Texas requires a minimum score on the direct writing section to pass the English language arts test.

The WASL includes two prompts, and the test-taker must respond to both. This approach is similar to New Jersey's. At the 10th grade level, one prompt is always expository and the other is persuasive — both of which are among the forms of writing that colleges and employers say high school graduates need to master. Indeed, Achieve's American Diploma Project (ADP) English benchmarks clearly call for the same kinds of writing that are expected in the WASL, stressing the importance of being able to develop a clear thesis; structure ideas in a sustained and logical fashion; support an argument with relevant details; and provide a coherent conclusion.

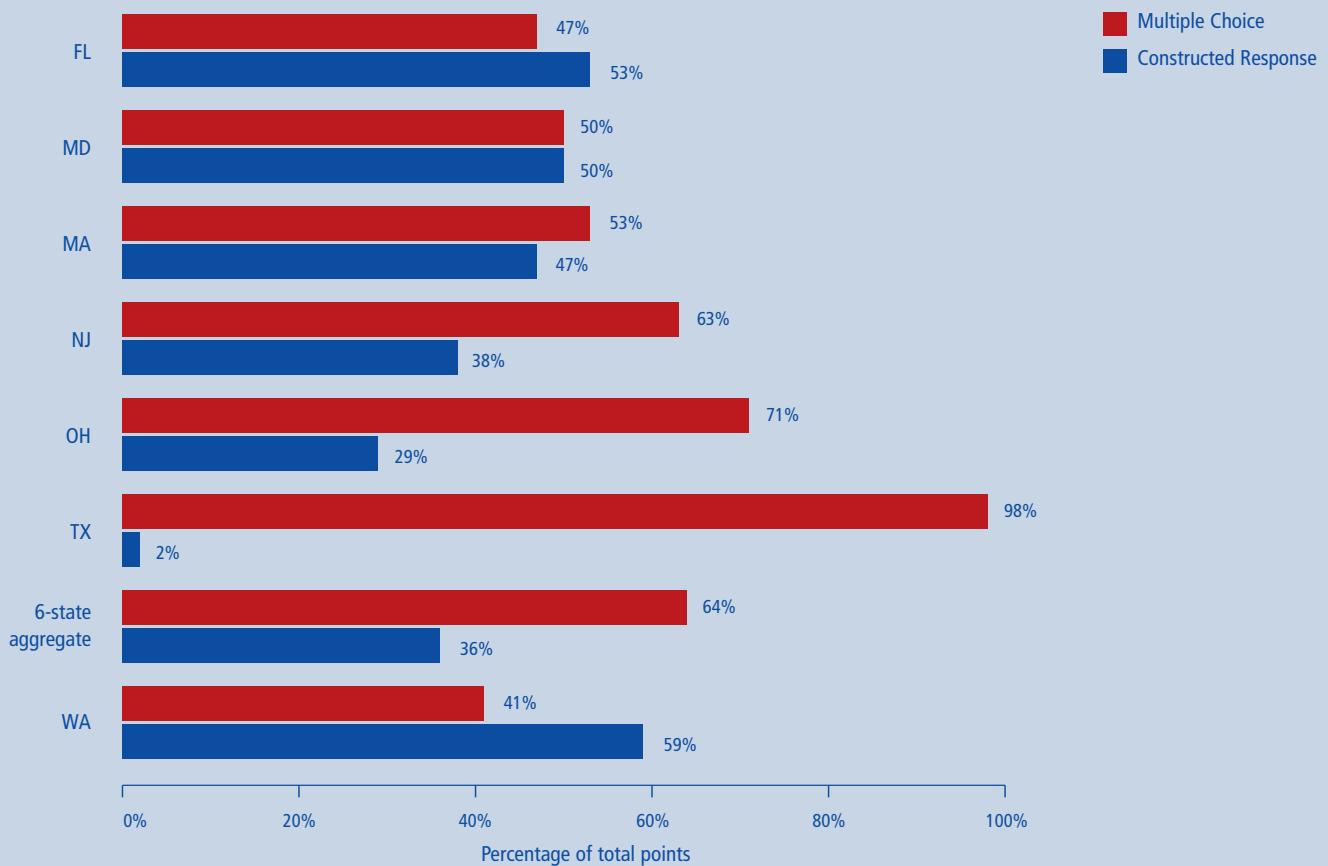
Although the rubric used for scoring the writing samples is not specific to a particular genre, the checklist included for the student writer in the test itself notes specific requirements of the genre and very clearly mirrors the expectations described in the ADP Benchmarks.

Mathematics

The WASL in mathematics includes more constructed-response items than other states' tests.

The WASL derives 59 percent of its points from constructed-response items and 41 percent from multiple-choice. Indeed, of all the tests examined, the WASL has the highest proportion of its points attributable to constructed-response items. This is notable, as constructed response items are often associated with tasks that require more cognitively challenging knowledge and skills and may require students to solve multi-step problems.

Chart 9: Distribution of points by item type



NOTE: *Totals may not equal 100 percent due to rounding.*

The WASL gives greater emphasis to number and data — and less emphasis to algebra and geometry — than other states' tests.

In our initial study of the six states, when Achieve divided the questions on the mathematics tests into the discipline's four major domains — number, algebra, geometry/measurement and data — we found that the majority (69 percent) of the points students could earn focused on algebra and geometry/measurement (31 percent and 38 percent respectively), followed by data (19 percent) and number (12 percent).

Chart 10: Distribution of points by content

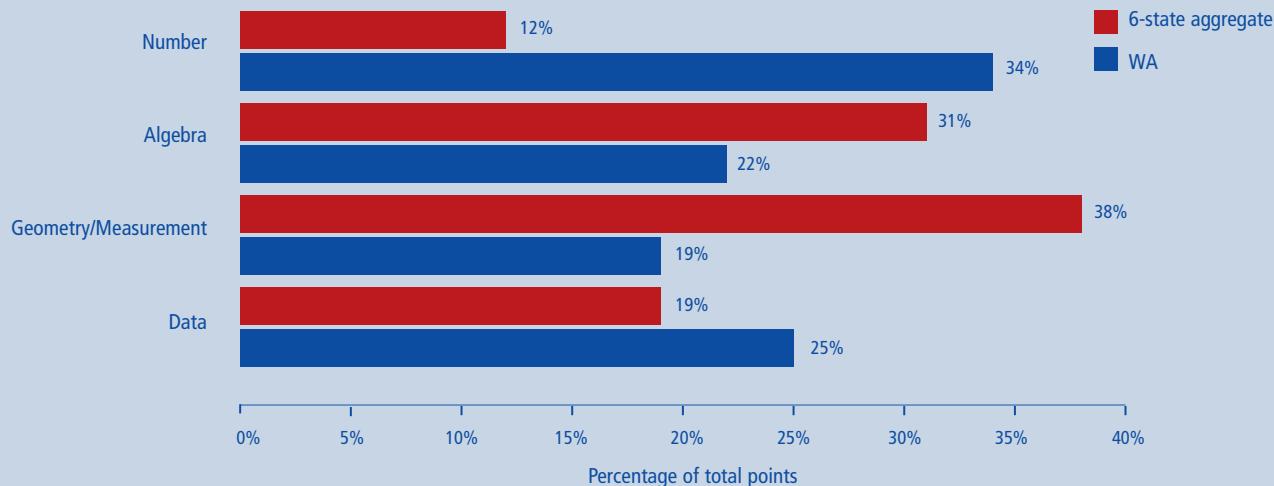


Table 3: Distribution of points by content

Discipline	FL	MD	MA	NJ	OH	TX	6-state aggregate	WA
Number	20%	3%	13%	27%	11%	8%	12%	34%
Algebra	25%	25%	40%	23%	27%	48%	31%	22%
Geometry/Measurement	40%	49%	27%	25%	40%	37%	38%	19%
Data	15%	23%	20%	25%	22%	7%	19%	25%

In contrast, the WASL devotes only 41 percent of its points to algebra and geometry/measurement (22 percent and 19 percent respectively), 34 percent to number, and 25 percent to data. While the WASL is comparable to 4 of the 6 states with respect to the degree of emphasis placed on algebra, it places much less emphasis on geometry.

The most significant difference between the WASL and the other tests is the emphasis it gives to number: Thirty-four percent of WASL points are attributed to number concepts, as compared with 12 percent of points across the other states' tests. Because number topics tend to be covered at earlier grade levels, it is not surprising that the WASL's level of rigor is lower than that of the other tests. The rigor of the WASL is further undermined by the fact that the items that address number concepts tend to focus predominantly on the least challenging aspects of number, such as whole number concepts, fractions, decimals and percents — topics that are typically addressed in middle school.

Table 4: Distribution of points by content: Number

Content Area	FL	MD	MA	NJ	OH	TX	WA	
Discrete Mathematics	8%	0%	13%	15%	20%	0%	14%	
Estimation	8%	0%	13%	8%	20%	0%	9%	
Fractions, Decimals, Percents	33%	67%	50%	31%	60%	60%	23%	
Number Theory	0%	0%	0%	0%	0%	0%	0%	
Whole Number Meaning, Operations and Properties	42%	0%	13%	31%	0%	20%	41%	
Basic Proportionality Concepts and Problems	8%	33%	13%	15%	0%	20%	14%	

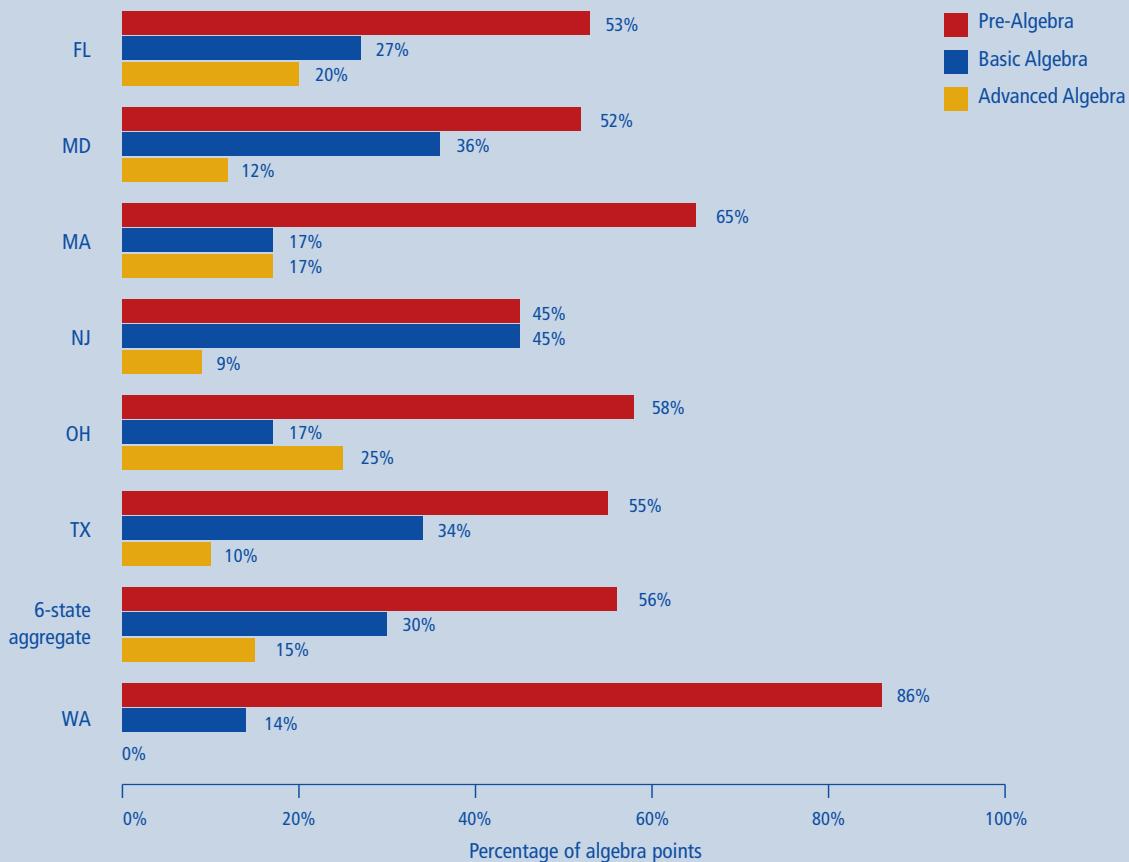
NOTE: *Totals may not equal 100 percent due to rounding.*

The WASL emphasizes pre-algebra over more advanced algebra.

Because algebra and geometry are such important topics, we took a closer look at the particular algebra and geometry/measurement topics being assessed. Across the six states in the original study we found that a majority of the algebra points students can earn are associated with the least demanding topics. Five of the six states have a majority of their algebra points assessing pre-algebra concepts that students should have mastered prior to high school. These include such basic skills as working with integers, rational numbers, patterns, representation, substitution, basic manipulation and simplification. In these six states, less than one-third of the points are dedicated to concepts such as linear equations, basic relations and functions typically associated with basic algebra or Algebra I — a course commonly taken in the ninth grade or earlier. An even smaller proportion of the algebra points (15 percent) reflect advanced algebra concepts typically encountered in Algebra II — or advanced algebra — courses. Few of the test questions measure skills college-bound students will need to succeed in credit-bearing college mathematics courses.

Compared with the other six states, the WASL has a significantly greater proportion (86 percent) of its algebra points attributable to items that assess prealgebra concepts. Only 14 percent of the algebra points on the WASL are attributable to items that assess basic algebra skills, compared with an average of 30 percent across the other six states. The WASL is the only one of the tests examined that does not assess advanced algebra at all. This emphasis on prealgebra and the lack of advanced algebra is another factor that lowers the rigor of the WASL overall.

Chart 11: Distribution of points by content: Algebra



NOTE: *Totals may not equal 100 percent due to rounding.*

The WASL is similar to the other six states' tests in its focus on two-dimensional geometry.

In our original study, Achieve found that half the geometry/measurement points on the six state tests were associated with two-dimensional geometry and measurement, while only a small proportion of the points (14 percent) were attributed to three-dimensional geometry — concepts such as volume and surface area. On the WASL, 58 percent of the geometry points assess two-dimensional aspects and 25 percent assess three-dimensional geometry. Geometry tends to be less hierarchical than algebra, so two-dimensional geometry is not necessarily less challenging than three-dimensional geometry. It is worth noting, however, that the National Assessment of Educational Progress (NAEP) includes two-dimensional geometry and measurement on its 8th grade assessment and expands to include formal three-dimensional geometry on its 12th grade assessment, indicating that it is considered to be end of high school level content.

Table 5: Distribution of points by content: Geometry/Measurement

Content Area	FL	MD	MA	NJ	OH	TX	6-state aggregate	WA
Congruence, Similarity, Transformations	29%	36%	25%	17%	28%	18%	28%	17%
2D Geometry and Measurement	42%	48%	44%	83%	44%	50%	49%	58%
3D Geometry and Measurement	13%	6%	31%	0%	17%	27%	14%	25%
Basic Measurement	17%	2%	0%	0%	6%	0%	4%	0%
Trigonometry	0%	8%	0%	0%	6%	5%	4%	0%

NOTE: *Totals may not equal 100 percent due to rounding.*

Like other state exit exams, the WASL measures mathematics concepts students in other countries study prior to high school.

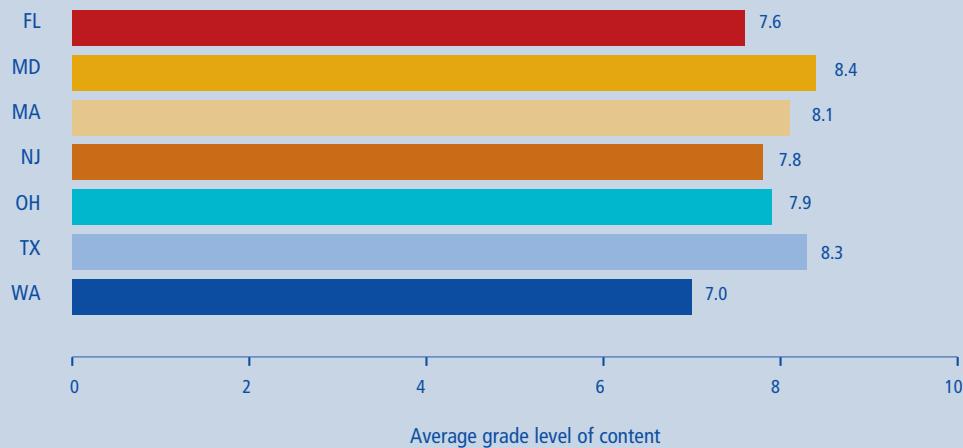
Because the performance of U.S. high school students in mathematics lags behind that of students in other industrialized countries, it is valuable to compare what is expected of students on these tests with expectations in other countries. In our exit exam study, Achieve had the advantage of looking at the mathematics exams by means of the International Grade Placement (IGP) index developed by Michigan State University as part of its ongoing work on the Third International Mathematics and Science Study (TIMSS).

The IGP index represents an “average” or composite among 41 nations of the world (both high-performing and low-performing countries) as to the grade level in which a mathematics topic typically appears in the curriculum. For example, decimals and fractions tend to be focused on at the 4th grade level internationally. Therefore, this topic has an IGP rating of 4. Right triangle trigonometry, on the other hand, is most often taught in the 9th grade around the world, so it receives an IGP rating of 9.

When applied to assessment items, the IGP describes content only. It is not intended to reflect performance demands (which are captured by another dimension of our methodology) or item format. When Achieve applied the IGP index to the six state exit exams, it revealed that the content measured on the tests is focused, on average, at the 8th grade level internationally. In other words, the material on the exams states are using as a requirement for high school graduation is considered middle school content in most other countries. While there was some variation across the states, no test had an average IGP rating higher than the eighth grade. The range of average IGP values across the six tests in the original study extended from a low of 7.6 for Florida to a high of 8.4 for Maryland.

As the following bar chart demonstrates, the average IGP value for the WASL is lower than those for the six state tests previously examined. In part this may be due to the emphasis on number, many aspects of which tend to fall fairly low on the IGP scale. In addition, as previously stated, the algebra items on the WASL tend to measure the lower level topics within that strand.

Chart 12: Test content on the International Grade Placement Scale



The majority of points on the WASL mathematics test are attributable to items that are at the middle to lower end of the cognitive continuum.

The content measured by the test items tells an important part of the story, but a more complete understanding of what these tests measure requires an examination of the cognitive demand of the items as well. In other words, what are students actually required to do with the content? Are students asked to apply routine procedures to mathematical problems? For example, does an item simply ask students to multiply two fractions to arrive at the answer? Or is the item framed in such a way that it requires students to first develop a more complex mathematical model to solve the problem? Essentially, the scale Achieve used to measure cognitive demand was designed to capture the *processes* that students employ as they “do” mathematics.

In our original study, Achieve found that a majority of the points on the tests across the six states are associated with items that require students to employ processes at the lower end of the cognitive continuum. On a five-point scale of rigor, with one being the least demanding and five being the most demanding, more than half the points across the tests are tied to the lowest two levels. An average of 48 percent of points across the six state mathematics tests are devoted to Level 2 items — items that require students to use routine procedures and tools to solve mathematics problems. About a quarter of the points across all of the tests are attributed to items that require more advanced mathematical skills (Levels 4 and 5).

The WASL follows this same general pattern, with 53 percent of its points attributable to items that call for either recall (Level 1) or the use of routine procedures (Level 2). As is true across all of the states, the bulk of the less cognitively demanding items are Level 2.

Chart 13: Distribution of points by level of cognitive demand

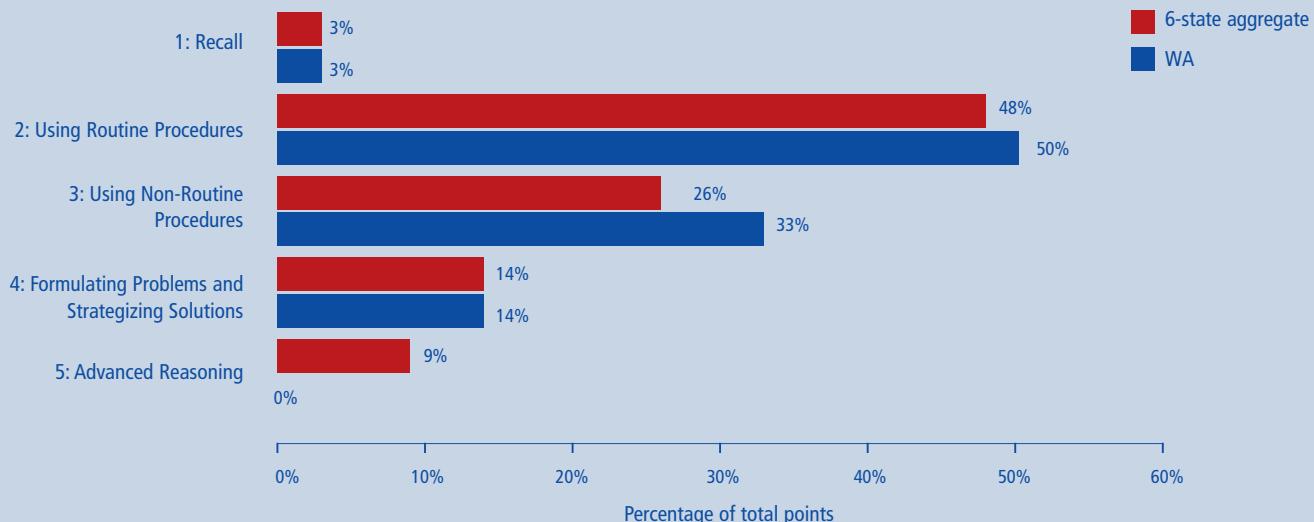


Table 6: Distribution of points by level of cognitive demand

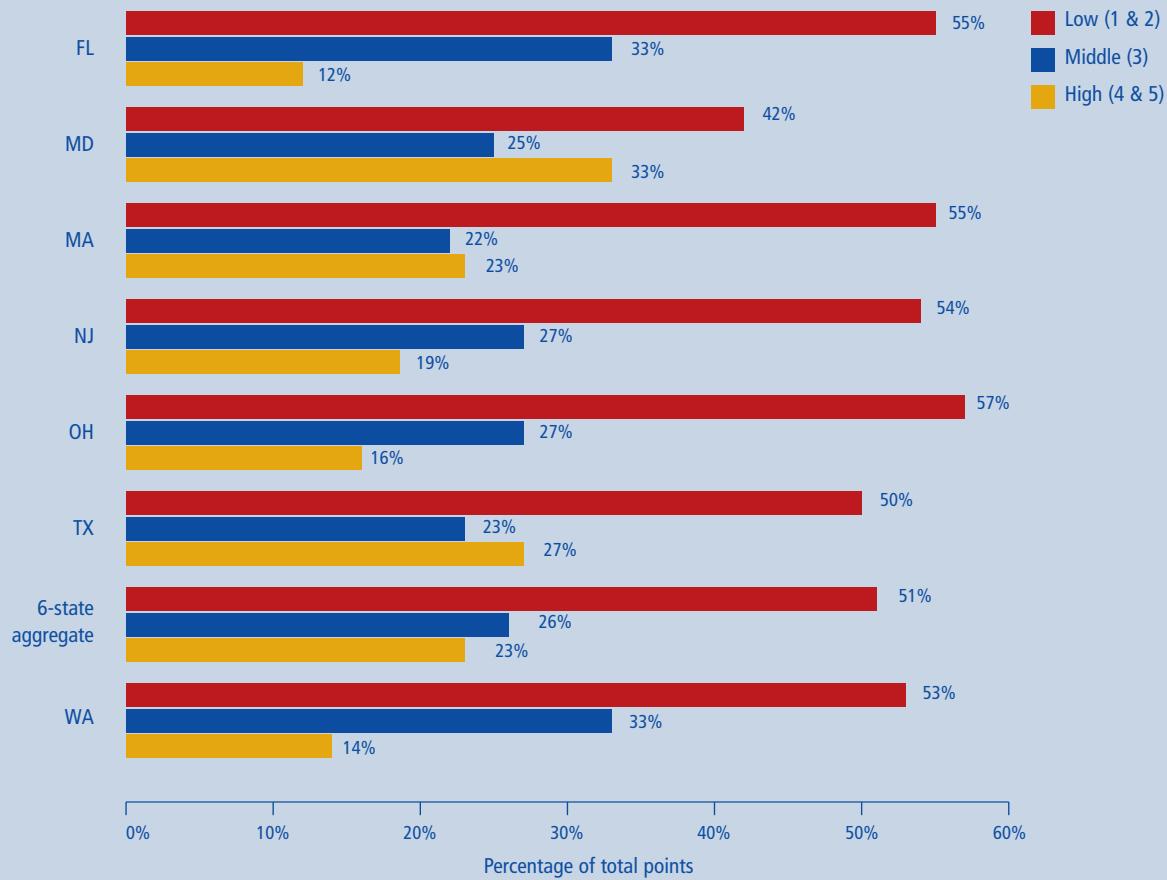
Cognitive Demand Level	FL	MD	MA	NJ	OH	TX	6-state aggregate	WA
1: Recall	2%	3%	2%	8%	4%	2%	3%	3%
2: Using Routine Procedures	53%	39%	53%	46%	53%	48%	48%	50%
3: Using Non-Routine Procedures	33%	25%	22%	27%	27%	23%	26%	33%
4: Formulating Problems and Strategizing Solutions	5%	17%	8%	15%	16%	22%	14%	14%
5: Advanced Reasoning	7%	17%	15%	4%	0%	5%	9%	0%

When compared with the other state tests examined, the WASL tends to have a relatively high proportion of its points (33 percent) attributed to items that ask students to use non-routine procedures (Level 3). Such procedures include estimating, comparing, classifying and using data to answer a question, and using mathematics to make decisions that go beyond a routine problem-solving activity. Only Florida — with 33 percent of its points attributable to Level 3 items — matches the WASL in this regard.

Similar to the other six states, the WASL places the least emphasis on the highest levels of cognitive demand: Only 14 percent of the WASL's points are attributable to Level 4 items that require students to formulate a problem, to strategize, or to critique a solution method. None of the WASL's points correspond to Level 5 items, which ask students to develop algo-

rithms, generalizations, conjectures, justifications, or proofs. Although some constructed response items may appear to require some of these processes, the scoring guides for these items indicate that points are awarded for performances associated with the lower levels of cognitive demand.

Chart 14: Distribution of points by level of cognitive demand



II. How do the performance levels on the WASL compare with those of other states?

The aim of a standards-based education system is for all students to acquire the knowledge and skills described by a state's content standards. State assessments are the principal tool for measuring how well students have mastered that content. Up until this point, this report has focused on what is measured on the WASL and six other state exit exams — the content, the difficulty level of the questions and the complexity of the reading passages. However, students taking these tests are not required to answer all of the questions correctly to pass. States establish “cut scores” that students need to achieve to pass the tests. These cut scores define the level of achievement that students are ultimately held accountable for — they establish the “floor” of performance required to earn a high school diploma. As such, these scores represent the level of mastery that a state deems satisfactory.

The Accountability Commission asked Achieve to compare the “Basic” and “Proficient” levels on the WASL with the cut scores students must reach to pass the tests in the other six states. The Commission used that comparative information to inform its fall 2004 decision regarding where to set the passing score on each test and how to set policy for the WASL graduation requirement.

Methodology

Comparative studies of where states set their cut scores are rare and difficult to conduct. They typically involve comparing the percentage of students passing each state's test with the percentage of students passing a common test, such as NAEP. This methodology permits judgments about the relative difficulty of different tests, but doesn't provide information on the knowledge and skills students need to pass each test.

Achieve, working with researchers from Michigan State University, developed a new procedure for comparing cut scores across state tests that focuses on the content of the test questions, thus giving states a clearer comparative picture of their expectations for students. The procedure was first used in *Do Graduation Tests Measure Up?*, published in June of 2004, and has been replicated here for the WASL analysis. Because the items on the WASL and the six other state tests have been coded according to common metrics discussed in the previous section of the report (e.g., content and cognitive demand), it is possible to use these metrics to identify what a typical student passing the assessments is expected to know and do. (For more information on this methodology, please refer to *Do Graduation Tests Measure Up?*.)

It is important to note that our initial six-state study focused only on students scoring at the passing score, not on those who scored at higher or lower levels on the tests. In this study of the WASL, we examined the cut scores that are used to determine the Basic and Proficient performance levels. We were able to compare what it takes to reach the Basic and Proficient levels on the WASL with what other states require students to do to “pass” their tests.

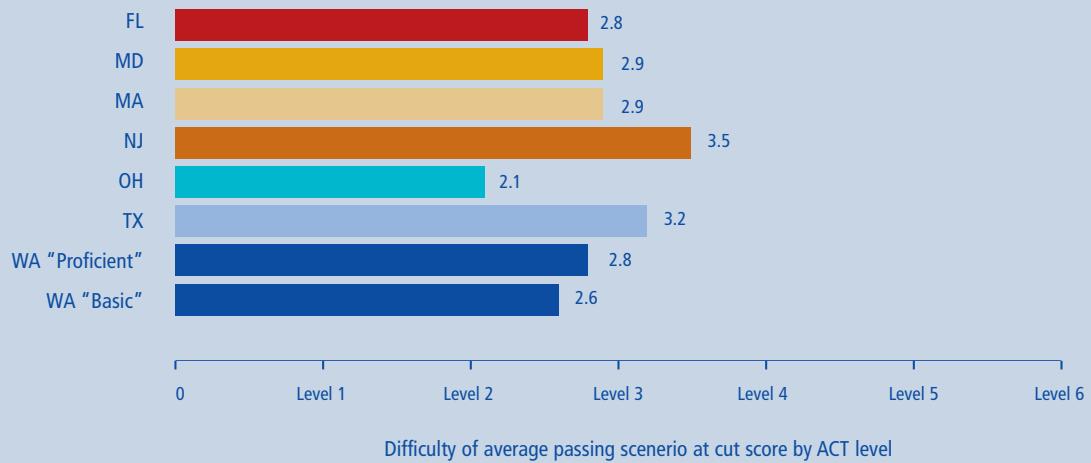
Performance Levels on the Reading Test

Achieve compared cut scores across the English language arts tests using the ACT skills hierarchy. As stated earlier, levels 1-3 are most heavily assessed on the ACT’s EXPLORE test, which is given to 8th and 9th graders. ACT’s PLAN test, given to 10th graders, focuses most heavily on Level 3-5 questions, while the college admissions exam — the ACT Assessment — focuses on Levels 4-6.

Given this frame, Achieve found that the average ACT skill level at the passing score on the state exit exams in the original study ranged from 2.1 to 3.5. Thus, students scoring at the passing level are, generally speaking, being asked to perform at the level that ACT considers appropriate for 8th and 9th graders.

Similarly, the average ACT skill level at the Proficient score (400) on the WASL reading test is 2.78, indicating that as in other states, students reaching Proficient on the WASL must exhibit the knowledge and skills that ACT treats as 8th and 9th grade content. The average across the six other states is 2.89, which suggests that scoring Proficient on the WASL is slightly less challenging than passing the exit exams in Massachusetts, Maryland and Florida. With average ACT skill levels of 3.47 and 3.19 respectively, the New Jersey and Texas tests appear to be the most challenging ones to pass among the seven, which is not surprising given the relatively high level of content and cognitive demand in these tests. (Note: Item format is not considered as part of this scale.) It also is worth noting that New Jersey and Texas administer their tests in the 11th grade, whereas most of the other states, including Washington, administer their tests in 10th grade. The exception is Maryland, whose end-of-course test is administered at the end of the 9th grade.

Chart 15: Difficulty of average passing scenario at “passing” cut score (ACT scale)



At the WASL's Basic level (375), the average ACT level is 2.6, which is lower than the average ACT level at the passing cut score on all other states' tests except for that of Ohio. On the WASL, to reach the Basic level students must exhibit, on average, slightly lower level knowledge and skills than those at the Proficient cut score (400) must, and must get fewer items correct to pass.

Performance Levels on the Writing Test

As stated earlier, Washington is the only state among the seven we discuss in this review to directly assess writing through a separate writing test that students are required to pass for graduation. Four of the six states in our initial study assess writing as part of the language arts assessment. None of these states render independent scores for writing, and Texas is the only state that requires students to pass the writing portion of the ELA assessment to pass the test as a whole.

Because the other states combine the writing and reading scores into their performance levels, we cannot compare the WASL writing cut score with that of other states. We can, however, comment on the scoring of the WASL.

In any discussion of cut scores for a writing assessment, one must recognize that a writing score is an indication of degrees of writing quality, not a representation of the number of items answered correctly, as in a reading test. Washington uses a six-point scale to judge the quality of writing tasks. Typically, a score of 4 or above describes an adequate performance for the grade level (as seen, for example, in the SAT Writing rubric that describes a 4 as “competent, adequate mastery” and a 3 as “inadequate, developing mastery”*).

On the 10th grade WASL, each student produces two writing samples, each of which is scored twice. The total possible score for writing at this level is 24, 12 possible points per sample.

The cut score for reaching level 3, the Proficient level, requires a total of 17 out of 24 points. While there are many possible combinations of points that a student could earn for this 17 point total, a typical score set would be 8 points for one sample (two scores of 4, a proficient level), and 9 points for the other (a score of 5 and a score of 4, a clearly proficient score). A score of 17, then, is an indication that the writer is competent for the grade level.

A level 2 performance, the Basic level, requires a total of 13 out of 24 possible points. Such a score would be most often achieved by receiving scores of 6 and 7. The 6 score indicates two scores of 3 for one sample and scores of 3 and 4 for the other. This reflects a minimally competent level of performance.

If a score of 15 were considered, the likely individual scores would be combinations of 8 (two scores of 4, indicating competency) and 7 (a score of 3 and 4, a performance very close to competent) for the two samples. A score of 15 would be an indication of fairly competent writing for the grade level.

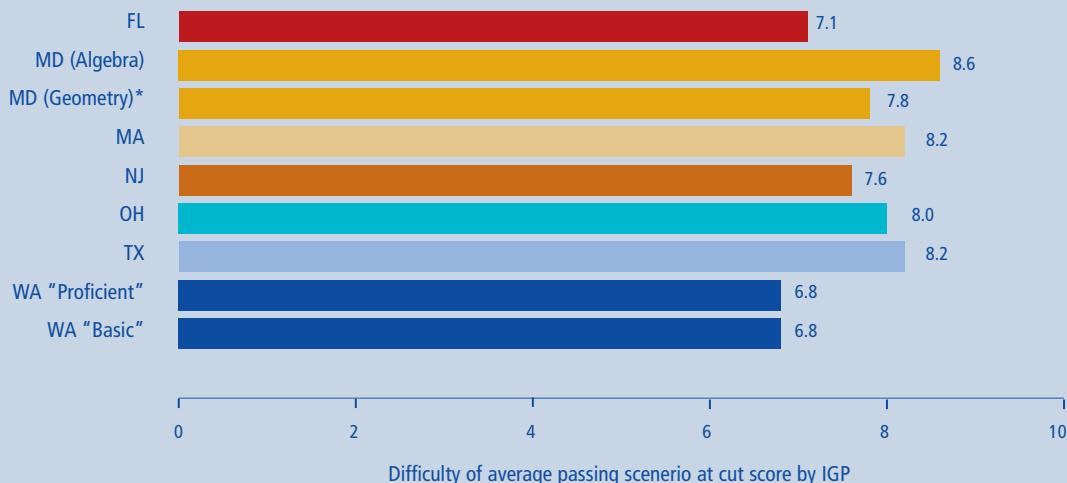
*Full rubric available at <http://www.collegeboard.com/student/testing/sat/about/sat/writing.html>

Performance Levels on the Mathematics Test

As described earlier, Achieve used the IGP index to identify the level of content measured on the tests. In our original study, we found that, on average, the tests from the six states measured mathematical content that tends to be focused on at the 8th grade level internationally. The level of mathematics content knowledge students need to pass the state exit exams ranged from 7.1 to 8.6. That is, the questions on the tests that students scoring at the cut score are likely to get correct measure, on average, concepts that students around the world focus on in the 7th and 8th grades. As Chart 16 indicates, Maryland's end-of-course algebra test appeared to be the most challenging one to pass in terms of content difficulty, with an IGP of 8.6. The Texas, Massachusetts and Ohio tests followed.

The average IGP score at the WASL's Proficient cut score (400) is 6.8, suggesting that this test is less challenging to pass in terms of its content difficulty than the six other states' tests analyzed. Essentially, this means that to pass the WASL, students are required to know mathematics content that is taught, on average, in the late 6th grade or early 7th grade internationally. This is likely due to the emphasis on number, because many topics within this strand receive relatively low IGP ratings. In addition, as previously stated, a large majority of algebra items on the WASL assess pre-algebra concepts. Again, it is content — not cognitive demand or item format — that is the basis for the IGP index.

Chart 16: Average difficulty of mathematical content at "passing" cut score (International Grade Placement Scale)

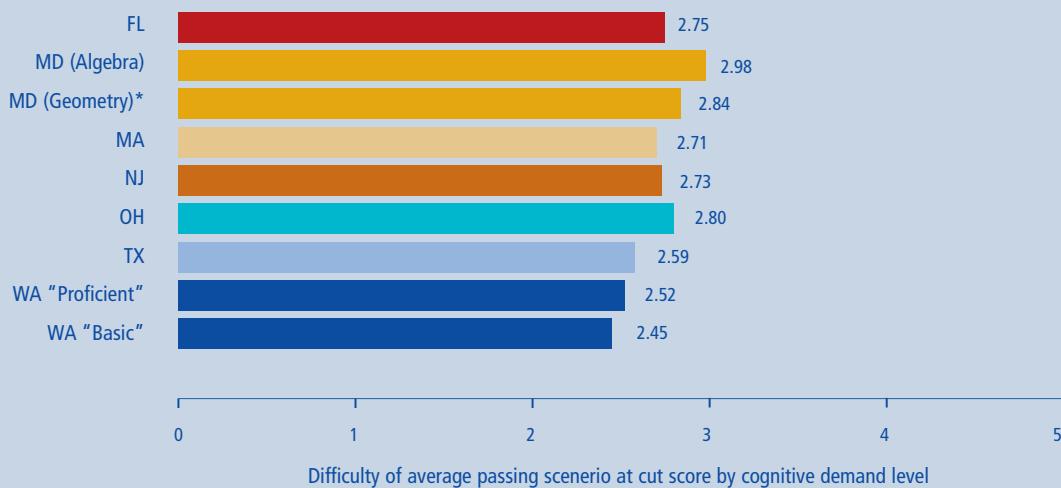


*Not required for graduation

Digging deeper, we examined the average cognitive demand of the items that students need to answer correctly to reach the Proficient level on the WASL and the passing levels on the other tests. Again, the WASL tends to fall at the lower end — 2.52 on a 5-point scale — in large part because 53 percent of the WASL's points are attributable to items that call for the lower levels of cognitive demands (Levels 1 and 2) and relatively few points (14 percent) are attributable to Levels 4 and 5, which require students to formulate a problem or to strategize or critique a solution method. In fact, zero points correspond to Level 5 items.

At the cut score for the WASL's Basic level (375), the IGP is virtually the same as it is at the Proficient cut score. However, the average cognitive load of the items at the Basic level (2.45) is slightly less challenging than at the Proficient level (2.52). This suggests that what differentiates students at the Proficient level from those at the Basic level is not necessarily a mastery of higher level content, but rather the ability to handle slightly more cognitively demanding items more consistently.

Chart 17: Average cognitive demand of mathematics tests at "passing" cut score



*Not required for graduation

Why aren't Washington students achieving higher scores in mathematics?

Despite the fact that the WASL appears on its face to be less challenging than most of the other mathematics tests Achieve analyzed, the fact remains that large numbers of students are not passing. In fact, although the average IGP at the WASL Proficient score is below that of the other six states, only 44 percent of Washington 10th graders passed the test in 2003 — the lowest passing rate among the states examined.

So what is it that makes the WASL mathematics test challenging for students? Are there characteristics of the test that could account for the relatively low scores that are not easily captured by our criteria in this study? In our judgment, there are two additional factors that may be contributing to low student performance: 1) a lack of motivation, as the test does not yet count for students; and 2) a lack of familiarity with the format of the test questions, which may be presenting greater challenges to students than we would expect.

There is growing evidence from other states that high school students take standards and assessments more seriously when they know their performance on those tests counts. For example, only 48 percent of 10th graders in Massachusetts passed the mathematics portion of the states' new graduation exam when it was first given in 1998. Some called for the state to lower the bar or delay implementation, but instead state officials and local educators redoubled their efforts to strengthen the curriculum and provide academic supports. When the 10th graders from the class of 2003 took the test — the first group that had to pass it to graduate — the scores jumped up nearly twenty percentage points, suggesting that when it counts, students (and schools) put forth more effort. By spring of 2003, 95 percent of students in the graduating class had passed the test.

A similar story played out in Virginia as it phased in new end-of-course exams for high school graduation. Only 40 percent of students passed the Algebra I exam when it was first given in 1998 (more students passed the reading and writing tests). By 2003, 78 percent had passed the Algebra I test, and by the time the first class of high school seniors had to pass several of the end-of-course tests to graduate in the spring of 2004, all but 1 percent earned their diplomas.

When we combine low student motivation with the significant role that open-ended and short-answer questions play in the WASL, this may begin to explain the low scores in mathematics. The WASL includes a large number of extended response and short answer tasks (59 percent of total points),

particularly when compared with other states (36 percent average across the other 6 states). These item types are essential because of their ability to measure more advanced skills and positively influence how material is taught in classrooms; however, if students have not had experience solving problems like these, their format can pose an additional challenge.

Even though in the case of the WASL the mathematical content of the items may not be as advanced as that on other state tests, the format of the questions may be challenging for students because there are not a set of answers to choose from. In addition, some of the items require a substantial amount of reading, and students often have to work through multiple steps to answer the question. It is possible that — because they know it doesn't count — students are not putting forth the necessary effort to complete the tasks.

To test this hypothesis, MSU researchers looked at the student-response data from the 2003 WASL in mathematics to see how students do on each of the questions. What they found was very revealing. Several of the short-answer items at the end of the test seem to be posing a particular challenge for students. On one of these items, 80 percent of a representative sample of students who took the test got zeros, suggesting that most students didn't even attempt to answer it. But of those who did answer the item, 75 percent earned the full 2 points, indicating that it was relatively easy for students if they ventured to try it. Furthermore, 82 percent of the students got the final test item — a multiple choice item — correct, indicating that they finished the test but may not have had the *motivation* to attempt a short answer item. We cannot know the mind-set of these students at the time they took the test, but it is conceivable that they were simply unmotivated to complete the items because the test does not count for them.

Conclusion

Achieve launched its original 2004 study to help answer some basic questions about the expectations states are setting for their high school graduates through the use of exit exams: Do the tests reflect material that students should be familiar with by the time they complete high school? Is it reasonable to expect all students to pass these tests before they graduate? If they pass these tests, does it mean students are ready for their next steps in life?

Across the states, we found that the tests do indeed set a floor for students that can be responsibly defended as a graduation requirement, but do not effectively tap the higher-level skills that truly constitute “readiness” for college and work.

In our analysis of Washington’s 2003 assessments, we found that compared with the other states, the writing test is exemplary. By requiring students to pass that test to graduate, Washington is placing more value on student writing than any of the other states in the study, which is commendable given how important strong writing skills are to students’ success in college and careers.

The WASL reading test is relatively strong as well. It includes challenging questions, although the reading passages are not as rigorous as in other states. The Proficient level of performance sets a standard that is comparable to other states in Achieve’s study.

The WASL mathematics test is the least challenging of the three when compared with the other states, most notably because the content is less rigorous. Given the relatively low level of content on the test, the Proficient level does not, in our opinion, set an unreasonable standard for high school graduates. The state is to be commended for including a large proportion of constructed-response items and contextualized multiple-choice items on the WASL.

• • • • •

In states such as Washington, where the exit exams are being debated, Achieve strongly encourages policymakers not to lower the standards or delay implementation. If Washington stays the course with these exams and makes the necessary investments to improve teaching and learning, it undoubtedly will find that students will rise to the challenge. When sufficient numbers of students pass these tests, Washington should continue to raise the floor to reflect the demands students will face in postsecondary education and the world of work.

Appendix: Summary of Methodology

To compare assessments, each assessment item was analyzed and coded according to a range of lenses designed to capture different characteristics of individual test items and the tests as a whole. Many of the criteria in English language arts and mathematics are similar, although there are important differences that stem from the distinct natures of the disciplines. To ensure the reliability of the data, at least two experts trained in the use of the criteria coded each test. Those experts reconciled any differences in coding before the data were analyzed.

The following are summaries of the various criteria according to which assessments in the study were analyzed. For the complete descriptions of the criteria, please visit Achieve's Web site at www.achieve.org.

Content of Items

Mathematics

This lens compares the content of state mathematics exams, using the Third International Mathematics and Science Study (TIMSS) Mathematics Framework adapted for use in this study by the U.S. TIMSS National Research Center at Michigan State University and Achieve experts. The framework provides a detailed, comprehensive taxonomy of mathematics content, organized at its most general levels according to the following major domains of mathematics:

- Number
- Algebra
- Geometry/Measurement
- Data

These domains are further broken down into smaller units to allow for finer-grained comparisons. For example, geometry content is divided into a variety of categories such as two-dimensional geometry and measurement; three-dimensional geometry and measurement; transformations, congruence and similarity; and trigonometry. The majority of these categories are subdivided even further to facilitate a high degree of content specificity in coding. Item coders for this study assigned up to three primary content codes to each test item. In many cases, the multiple content codes aligned with the same reporting category (e.g., geometry/measurement or algebra), but this was not always the case. Items that aligned with more than one reporting category were re-examined, and one primary code was identified.

English Language Arts

To identify the content on English language arts assessments, Achieve used a comprehensive taxonomy of the domains of reading, writing and language skills developed by the Council of Chief State School Officers (CCSSO) and adapted by Achieve experts. The CCSSO framework was developed in collaboration with several states that are a part of the Surveys of Enacted Curriculum.

Based on this framework, Achieve developed a taxonomy that included all the aspects of English language arts described in state standards — and therefore targeted in state tests — to describe as accurately as possible the content or topic that each item measured. The study required a taxonomy that was as specific as possible, providing sufficient discrimination among the topics to yield a clear portrait of what each state was emphasizing in its assessment of English language arts.

The major reporting codes for reading are:

- Basic comprehension (includes word definitions, main idea, theme and purpose)
- Literary topics (includes figurative language, poetic techniques, plot and character)
- Informational topics (includes structure, evidence and technical elements)
- Critical reading (includes appeals to authority, reason and emotion; validity and significance of assertion or argument; style in relation to purpose; and development and application of critical criteria)

The reporting categories for writing are:

- Writing (All items included in this category were direct writing assessments, typically writing in response to a prompt that asks students to address a particular question or thesis in a narrative, expository or persuasive essay. Although all such assessments included attention to language conventions, either as part of a holistic scale or as a discrete rubric, all direct writing tasks were coded to this category only and not coded as well to editing and revising or to grammar, mechanics and usage.)
- Editing and revising (Items coded to this category assessed the following topics through multiple-choice items: editing for conventions; organizing for meaning; and revising for meaning, style and voice.)
- Grammar, mechanics and usage (Items coded to this category assessed the following topics through multiple-choice items: spelling, mechanics, punctuation, syntax and sentence structure, grammatical analysis, and language usage.)

Approximate Grade-Level Demand of Items

Mathematics

To approximate the grade-level demand of mathematics items, Achieve used the TIMSS International Grade Placement (IGP) index, developed by the U.S. TIMSS National Research Center at Michigan State University. The IGP index represents a kind of composite among the 40 TIMSS countries (other than the United States) to show when the curriculum focuses on different mathematics content — at what point the highest concentration of instruction on a topic occurs. Using their nation's content standards document, education ministry officials and curriculum specialists in each TIMSS country identified the grade level at which a mathematics topic is introduced into the curriculum, focused on and completed. The IGP index is a weighted average of those determinations. For example, a topic with an IGP of 8.7 is typically covered internationally toward the end of 8th grade. The content topics to which Achieve coded test items all have an IGP value associated with them. For items that spanned more than one category and were subsequently assigned a single code, the retained content code tended to be that with the highest IGP value.

The following are examples of the IGP ratings of various mathematics topics.

CONTENT DESCRIPTION	IGP INDEX
Whole Number: Operations	2.5
Rounding and Significant Figures	4.7
Properties of Common and Decimal Fractions	5.6
Exponents, Roots and Radicals	7.5
Complex Numbers and Their Properties	10.7

English Language Arts

To approximate the grade level demand of English language arts items, Achieve adapted the ACT Standards for Transition (for English language arts and reading), which provide a hierarchy of skills in these topic areas by taking into account the performance and content of an item as well as the reading demand of the reading passage being assessed. ACT, Inc.'s Educational Planning and Assessment System encompasses three assessments administered during 8th and 9th grades, 10th grade, and 11th and 12th grades. The Standards for Transition form the basis of all three, with each successive test including more complex content and performances from the standards. The standards are divided into six levels:

- Levels 1 through 4 are assessed on the EXPLORE test (8th and 9th grades);
- Levels 1 through 5 are assessed on the PLAN test (10th grade); and
- Levels 1 through 6 are assessed on the ACT Assessment (11th and 12th grades).

STANDARD	COMPARATIVE RELATIONSHIPS
LEVEL 4	Have a sound grasp of relationships between people and ideas in uncomplicated passages.
LEVEL 5	Identify clearly established relationships between characters and ideas in more challenging literary narratives.
LEVEL 6	Reveal an understanding of the dynamics between people and ideas in more challenging passages.
	Make comparisons, conclusions and generalizations that reveal a feeling for the subtleties in relationships between people and ideas in virtually any passage.

Cognitive Demand of Items

Mathematics

This lens provides a taxonomy of performance expectations (what students are expected to “do” with the mathematics content they know) based on a synthesis of the TIMSS Mathematics Framework and Achieve’s assessment-to-standards alignment work with states. The five-point scale provides information on the kind and complexity of performance required of students — ranging from simple recall of information to complex reasoning skills.

- Level 1 includes demonstrating basic knowledge or recall of a fact or property.
- Level 2 includes routine problem-solving that asks students to do such things as compute, graph, measure or apply a mathematical transformation.
- Level 3 includes estimating, comparing, classifying and using data to answer a question or requiring students to make decisions that go beyond a routine problem-solving activity.
- Level 4 includes asking students to formulate a problem or to strategize or critique a solution method.
- Level 5 includes asking students to develop algorithms, generalizations, conjectures, justifications or proofs.

Coders often assigned multiple performance codes to items. Sometimes primary performance codes for an item spanned two or more of the reporting levels. In cases such as this, each item was re-examined, and a decision rule was made to accept the highest performance level category as representing the performance expectation of that item.

English Language Arts

The cognitive demand lens for English language arts provides a taxonomy of performance expectations based on Achieve’s assessments-to-standards alignment protocol and CCSSO’s description of performances in its Survey of Enacted Curriculum. Four levels of reading cognitive complexity provide information on the kind and complexity of reasoning required of students, ranging from simple recall of information to complex reasoning skills.

- Level 1, under the heading “Literal Recall,” covers such skills as providing facts, terms and definitions; describing ideas; locating answers in a text; identifying relevant information; and identifying grammatical elements.
- Level 2, under the heading “Infer,” covers such skills as inferring from local data, inferring from global data, drawing conclusions, identifying purposes, identifying main ideas or theme, identifying organizational patterns, and predicting.
- Level 3, under the heading “Explain,” includes such skills as following directions, giving examples, summarizing information, checking consistency and recognizing relationships.
- Level 4, under the heading “Analyze,” covers such skills as categorizing; distinguishing fact from opinion; ordering, grouping, outlining and organizing ideas; comparing and contrasting ideas; and interpreting techniques.

Demand of Reading Passages

Achieve analyzed the difficulty level of each reading passage according to a six-point scale ranging from straightforward text to more complex, challenging and abstract text. This scale was developed by noted reading experts who reviewed various characteristics of passages, such as level or specialization of vocabulary, predictability of structures or organization, complexity of syntax, level of abstractness, familiarity of the topic, and the number of concepts introduced in the passage. Generally speaking, Level 1 represents upper-elementary reading levels, Levels 2 and 3 represent middle school-level reading, Level 4 represents early-stage high school reading, and Levels 5 and 6 represent late-stage high school reading.

Categories for consideration of reading passage difficulty include:

- **Structure**
 - Narration
 - Description
 - Explanation
 - Instruction
 - Argumentation
- **Vocabulary**
 - Poetic
 - Idiomatic
 - Technical
 - Unusual/unfamiliar
- **Syntax/connectives**
 - Dialogue
 - Sentence structure
- **Characters/ideas**
- **Narrator/stance**
- **Theme/message/moral**
- **Literary effects**
 - Foreshadowing
 - Flashback
 - Irony
- **Familiarity**
 - Topic
 - Place
 - Time period

(For examples of reading passages at all six levels, please visit Achieve's Web site at www.achieve.org.)

Reading Rigor Index

The Reading Rigor Index (RRI) is a method of determining how the cognitive demand of an item interacts with the level of a reading passage. For example, an item could require a low performance of a difficult passage, a high performance of an easy passage, a high performance of a difficult passage or a low performance of an easy passage. This interaction of level of cognitive demand and reading level contributes to the challenge of an item. Items also are accounted varying point values. An item attributed one point is weighted less than an item attributed two or more points.

The RRI score is obtained by adding the cognitive demand level and the reading demand level for each reading item on a test. The Cognitive Demand Scale ranges from a low of one to a high of four and the Reading Level Demand Scale from a low of one to a high of six. This makes nine Reading Rigor levels — Level 1 for items with the lowest possible score of two (an item with a cognitive demand of one and a reading level demand of one); Level 9 for the highest possible score of 10 (an item with a cognitive demand level of four and a reading demand level of six).

An item's point value determines the number of times the RRI score is counted to determine RRI percentages. An item worth two points will have its RRI score counted twice, an item worth three points will have its score counted three times and so on.

“Cut Scores”

Each state determines the levels of proficiency its students must reach to pass the state's exit exam based on scaled scores. The difficulty in comparing performance levels and the cut scores that reveal these levels is that these scaled scores are unique to each state's exam and students. Without a comparison sample — giving different state exams to the same group of students or giving a common exam to students in all six states — no connections among these scaled score distributions exist. Consequently, aside from a subjective analysis of proficiency-level setting procedures, it has been impossible to determine objectively if the proficiency levels set by different states have similar meaning.

Achieve, working with researchers from Michigan State University, developed a procedure to establish comparability of proficiency levels across states according to the different dimensions by which the assessments analyzed in this study have been coded. Because the assessments from the six states were coded item by item according to common metrics, it became possible to compare what passing the assessments exactly at the cut score would mean, state to state. Achieve chose, in this study, to look at the mathematics cut scores through the IGP index lens and the English language arts cut scores through the ACT index (both are described above).

States almost universally use Item Response Theory (IRT) models to scale assessment items and to estimate a scaled value for each student. The cut score is established in this metric. Consequently, the cut scores (the scores needed simply to pass, not reach any level of greater proficiency) and scaling information provided by the states were used to determine sets of correctly answered items — or passing “scenarios” — that allow students to reach the cut score and the likelihood that those scenarios would occur. When coupled with the IGP (for mathematics) or ACT (for English language arts) codings of the items, the process transforms the cut scores into the corresponding IGP or ACT metrics. Comparisons of states' cut scores are done in these metrics. Because of the large number of potential passing scenarios (2^n where n is the number of items or points on the test), only a random sample of 20,000 passing scenarios were used for the computation.

Acknowledgments

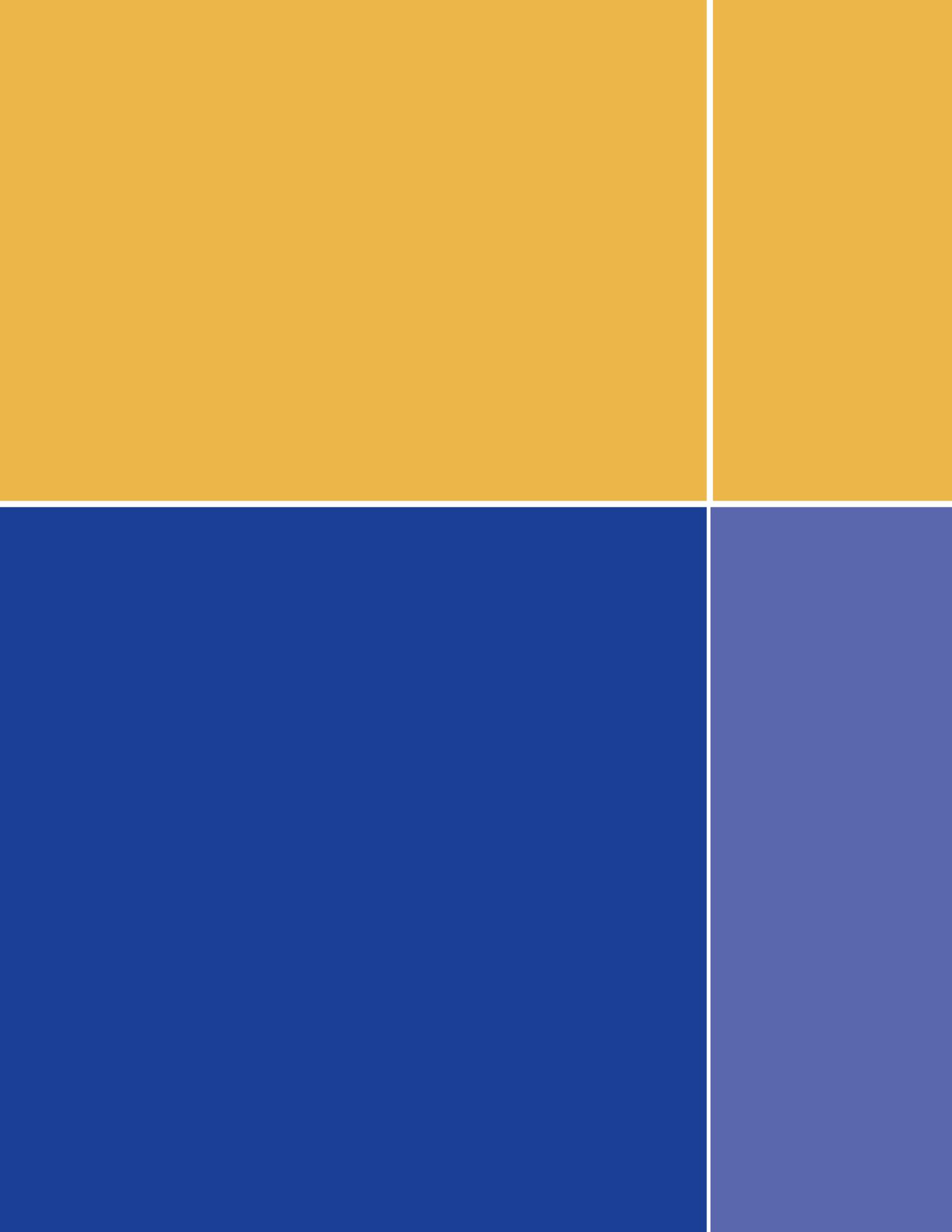
Achieve would like to thank the individuals and organizations whose contributions made this report possible.

Achieve senior associates JoAnne Eresh and Kaye Forgione led the analysis work in English language arts and mathematics, respectively. Jerome Halpern and Elizabeth Haydel, education consultants, provided content expertise in English language arts, as did Mara Clark, Achieve research associate, who also provided significant editorial assistance. Kathleen A. Wight, from the University of Michigan, provided content expertise in mathematics.

Matthew Gandal, Achieve's executive vice president, and Laura McGiffert Slover, director of the Mathematics Achievement Partnership, provided overall leadership and direction for the study.

The U.S. TIMSS National Research Center at Michigan State University provided considerable analytic and technical expertise. In particular, we would like to thank William H. Schmidt and Richard T. Houang for their substantial contributions.

Special thanks are also due to Kathy Ames, Jessica Palmer and Shona Stanish at KSA-Plus Communications, Inc., for their editorial advice and their help designing this report.





Achieve, Inc.
1775 Eye Street, NW
Suite 410
Washington, DC 20006
www.achieve.org